arXiv:1110.5281v2 [math.NA] 12 Jul 2012

MULTIGRID SOLUTION OF A DISTRIBUTED OPTIMAL CONTROL PROBLEM CONSTRAINED BY THE STOKES EQUATIONS

ANDREI DRĂGĂNESCU* AND ANA MARIA SOANE†

Abstract. In this work we construct multigrid preconditioners to accelerate the solution process of a linear-quadratic optimal control problem constrained by the Stokes system. The first order optimality conditions of the control problem form a linear system (the KKT system) connecting the state, adjoint, and control variables. Our approach is to eliminate the state and adjoint variables by essentially solving two Stokes systems, and to construct efficient multigrid preconditioners for the Schur-complement of the block associated with the state and adjoint variables. These multigrid preconditioners are shown to be of optimal order with respect to the convergence properties of the discrete methods used to solve the Stokes system. In particular, the number of conjugate gradient iterations is shown to decrease as the resolution increases, a feature shared by similar multigrid preconditioners for elliptic constrained optimal control problems.

Key words. multigrid methods, PDE-constrained optimization, Stokes equations, finite elements

AMS subject classifications. 65K10, 65N21, 65N55, 90C06

1. Introduction.

Over the last decade, the computational community has shown a growing interest in devising fast solution methods for large-scale distributed optimal control problems constrained by partial differential equations (PDEs). Optimal control problems constrained by the Stokes system form a stepping stone in the natural progression from the – now classical – Poisson-constrained test problem to problems constrained by more specialized and complex PDE systems modeling fluid flow such as Navier-Stokes, non-Newtonian Stokes, or the shallow water equations. Optimal control problems constrained by such PDE models play important roles in real-world applications, such as modeling of ice sheets or data assimilation for ocean flows and weather models.

In this article we consider an optimal control problem with a cost functional of tracking-type:

$$\min_{\vec{u},p,\vec{f}} J(\vec{u},p,\vec{f}) = \frac{\gamma_u}{2}\|\vec{u}-\vec{u}_d\|^2_{L_2(\Omega)^2} + \frac{\gamma_p}{2}\|p-p_d\|^2_{L_2(\Omega)} + \frac{\beta}{2}\|\vec{f}\|^2_{L_2(\Omega)^2}\ , \tag{1.1}$$

subject to the constraints

$$\begin{cases} -\Delta\vec{u} + \nabla p = \vec{f} & \text{in } \Omega,\\ \operatorname{div}\vec{u} = 0 & \text{in } \Omega,\\ \vec{u} = 0 & \text{on } \partial\Omega, \end{cases} \tag{1.2}$$

where Ω is a bounded polygonal domain in $\mathbb{R}^2$. The purpose of the control problem is to identify a force $\vec{f}$ that gives rise to a velocity $\vec{u}$ and/or pressure p to match a known target velocity $\vec{u}_d$, respectively pressure p_d. Since this problem is ill-posed,

[*] Department of Mathematics and Statistics, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, Maryland 21250 (`draga@umbc.edu`). The work of this author was supported in part by the Department of Energy under contract no. de-sc0005455, and by the National Science Foundation under award DMS-1016177.

[†] Department of Mathematics and Statistics, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, Maryland 21250 (`asoane@umbc.edu`). The work of this author was supported in part by the Department of Energy under contract no. de-sc0005455.

we consider a standard Tikhonov regularization for the force with the regularization parameter β being a fixed positive number. The constants γ_u, γ_p are nonnegative, not both zero.

The main goal of this article is to construct and analyze optimal order multigrid preconditioners to be used in the solution process of (1.1)-(1.2). Over the last few years a significant amount of work has been devoted to developing multigrid methods for optimal control problems. An overview of this research and further references can be found in a review article by Borzi and Schulz [1]. However, fewer works are dedicated specifically to optimal control problems constrained by the Stokes system. For example, recently, Rees and Wathen [11] have proposed two preconditioners for the optimality system in the distributed control of the Stokes system, a block-diagonal preconditioner for MINRES and a block-lower triangular preconditioner for a non-standard conjugate gradient method. We note that there are several papers in the literature on finite element error analysis for the optimal control of the Stokes equations (see, e.g., [10, 12, 4] and the references therein). Our paper focuses on the solution of the linear system that arises in the discretization process, which is not addressed in these papers. However, for completeness, we also prove an a priori error estimate for the optimal control since the cost functional in (1.1) includes a pressure term, using standard techniques similar to those used in [10, 12, 5].

Since the cost functional in (1.1) is quadratic, the KKT system is a linear saddle-point problem connecting the state, adjoint, and control variables. Solution methods for these problems typically fall into two categories: the all-at-once approach takes advantage of the sparsity of the system, but has the disadvantage that the matrix is indefinite. On the other hand, Schur-complement strategies may lead to smaller systems that may also be positive definite, but the sparsity is lost. Our approach is to eliminate the state and adjoint variables by essentially solving two Stokes systems using specific methods (see [7]), and to construct efficient multigrid preconditioners for the Schur-complement of the block associated to the state and adjoint variables. The constructed multigrid preconditioners are related to the ones developed by Drăgănescu and Dupont [5], and are shown to be of optimal order with respect to the convergence properties of the discrete finite element methods used to solve the Stokes system. In particular, the number of conjugate gradient iterations is shown to decrease as the resolution increases, a feature shared by similar multigrid preconditioners for elliptic-constrained optimal control problems. One word on the optimality of the preconditioner: the usual notion of optimality, especially in the context of multigrid, refers to the cost of the solution process being proportional to the number of variables. We argue that for this problem, an unpreconditioned application of conjugate gradient (CG) in conjunction with an optimal multigrid solve for the Stokes system already satisfies this notion of optimality. In the current context, multigrid preconditioners actually can perform better than that, and optimality refers to the order of approximation of the operator under scrutiny – in this case the reduced Hessian – by the multigrid preconditioner, as shown in Theorem 3.1.

The paper is organized as follows: In Section 2, we introduce the discrete optimal control problem and prove finite element estimates that will be needed for the multigrid analysis. Section 3 contains the main result of the paper, Theorem 3.1, which refers to the analysis of the two-grid preconditioner; furthermore, the extension to multigrid preconditioners is briefly discussed. In Section 4 we present numerical experiments that illustrate our theoretical results, and we formulate some conclusions in Section 5.

2. Discretization and convergence results. The strategy we adopt is to first discretize the optimal control problem then optimize the discrete problem. To define a discrete problem based on a finite element approximation of the Stokes system we briefly recall the usual weak formulation of the Stokes equations. Define the spaces

$$\begin{aligned} V &= H_0^1(\Omega)^2, \\ M &= L_{2,0}(\Omega) = \left\{ p \in L_2(\Omega) : \int_\Omega p = 0 \right\}, \end{aligned}$$

and the bilinear forms $a : V \times V \to \mathbb{R}$ and $b : V \times M \to \mathbb{R}$ as

$$\begin{aligned} a(\vec{u}, \vec{\varphi}) &= \sum_{i=1}^{2} \int_\Omega \nabla \vec{u}_i \cdot \nabla \vec{\varphi}_i, \\ b(\vec{\varphi}, p) &= -\int_\Omega p \operatorname{div} \vec{\varphi}. \end{aligned}$$

Throughout this paper, we write $(\cdot,\cdot)$ for the inner product in $L_2(\Omega)$ or $L_2(\Omega)^2$ according to context, and similarly for the norms, if there is no risk of misunderstanding.

The weak solution $(\vec{u}, p) \in V \times M$ of (1.2) is the solution of

$$\begin{aligned} a(\vec{u}, \vec{\varphi}) + b(\vec{\varphi}, p) &= (\vec{f}, \vec{\varphi}) && \forall \vec{\varphi} \in V, \\ b(\vec{u}, \psi) &= 0 && \forall \psi \in M. \end{aligned}$$

For $\vec{f} \in H^{-1}(\Omega)^2$ the problem has a unique solution [8]. Moreover, if Ω is a convex polygon and $\vec{f} \in L_2(\Omega)^2$, then $\vec{u} \in H^2(\Omega)^2$, $p \in H^1(\Omega)$ [9], and there exists $C = C(\Omega) > 0$ such that

$$\|\vec{u}\|_{H^2(\Omega)^2} + \|\nabla p\|_{L_2(\Omega)} \le C \|\vec{f}\|_{L_2(\Omega)^2}. \tag{2.1}$$

Throughout this paper we will assume Ω to be convex, so that the H^2- regularity of the Stokes problem is ensured. Furthermore, the target velocity field $\vec{u}_d$ is assumed to be from $L_2(\Omega)^2$ and the target pressure p_d from M.

We introduce the solution mappings $\mathcal{U}$ and $\mathcal{P}$ of the continuous state equation defined such that for any $\vec{f} \in L_2(\Omega)^2$ the following holds:

$$a(\mathcal{U}\vec{f}, \vec{\varphi}) + b(\vec{\varphi}, \mathcal{P}\vec{f}) = (\vec{f}, \vec{\varphi}) \quad \text{and} \quad b(\mathcal{U}\vec{f}, \psi) = 0 \quad \forall (\vec{\varphi}, \psi) \in V \times M.$$

The mapping $\mathcal{U}$, considered as a linear operator in $L_2(\Omega)^2$, is compact and self-adjoint, as

$$(\mathcal{U}\vec{f}_1, \vec{f}_2) = a(\mathcal{U}\vec{f}_1, \mathcal{U}\vec{f}_2) = (\vec{f}_1, \mathcal{U}\vec{f}_2) \quad \forall \vec{f}_1, \vec{f}_2 \in L_2(\Omega)^2.$$

We denote by $\mathcal{P}^* : L_{2,0}(\Omega) \to L_2(\Omega)^2$ the adjoint operator of $\mathcal{P}$, defined by

$$(\mathcal{P}^* q, \vec{f})_{L_2(\Omega)^2} = (q, \mathcal{P}\vec{f})_{L_2(\Omega)} \quad \forall q \in L_{2,0}(\Omega), \vec{f} \in L_2(\Omega)^2.$$

With this notation, the problem (1.1)-(1.2) is written in reduced, unconstrained form as

$$\min_{\vec{f} \in L_2(\Omega)^2} \hat{J}(\vec{f}) = \frac{\gamma_u}{2} \|\mathcal{U}\vec{f} - \vec{u}_d\|^2_{L_2(\Omega)^2} + \frac{\gamma_p}{2} \|\mathcal{P}\vec{f} - p_d\|^2_{L_2(\Omega)} + \frac{\beta}{2} \|\vec{f}\|^2_{L_2(\Omega)^2} \ . \tag{2.2}$$

The Hessian operator associated to the reduced cost functional $\hat{J}$ is given by

$$H_\beta = \gamma_u \mathcal{U}^* \mathcal{U} + \gamma_p \mathcal{P}^* \mathcal{P} + \beta I \ .$$

Note that the solution of the minimization problem (1.1) is obtained as the solution of the normal equation

$$H_\beta \vec{f} = \gamma_u \mathcal{U}^* \vec{u}_d + \gamma_p \mathcal{P}^* p_d. \tag{2.3}$$

The goal of this paper is to design an efficient multigrid algorithm for solving the discrete version of (2.3).

2.1. Finite element approximation. We consider a shape regular quasi-uniform quadrilateral mesh $\mathscr{T}_h$ of $\bar{\Omega}$, and we assume that the mesh $\mathscr{T}_h$ results from a coarser regular mesh $\mathscr{T}_{2h}$ from one uniform refinement. We use the Taylor-Hood $\mathbf{Q}_2 - \mathbf{Q}_1$ finite elements to discretize the state equation. The velocity field $\vec{u}$ is approximated in the space $V_h^0 = V_h \cap H_0^1(\Omega)^2$, where

$$V_h = \{v_h \in C(\bar{\Omega})^2 : v_h|_T \in Q_2(T)^2 \text{ for } T \in \mathscr{T}_h\},$$

and the pressure p is approximated in the space

$$M_h = \{q_h \in C(\Omega) \cap L_{2,0}(\Omega) : q_h|_T \in Q_1(T) \text{ for } T \in \mathscr{T}_h\},$$

where $Q_k(T)$ is the space of polynomials of degree less than or equal to k in each variable [3]. The control variable $\vec{f}$ is approximated by continuous piecewise biquadratic polynomial vectors in the space V_h.

REMARK 2.1. *For convenience, we choose here the quadrilateral $\mathbf{Q}_2 - \mathbf{Q}_1$ Taylor-Hood elements, however, our analysis can be extended to triangular $\mathbf{P}_2 - \mathbf{P}_1$ elements as well as other stable mixed finite elements.*

For a given control $\vec{f} \in L_2(\Omega)^2$, the solution $(\vec{u}_h, p_h) \in V_h^0 \times M_h$ of the discrete state equation is given by

$$\begin{aligned} a(\vec{u}_h, \vec{\varphi}_h) + b(\vec{\varphi}_h, p_h) &= (f, \vec{\varphi}_h) & \forall \vec{\varphi}_h \in V_h^0, \\ b(\vec{u}_h, \psi_h) &= 0 & \forall \psi_h \in M_h. \end{aligned}$$

Let $\mathcal{U}_h$ and $\mathcal{P}_h$ be the solution mappings of the discretized state equation and $\mathcal{U}_h^*$, $\mathcal{P}_h^*$ their adjoints, defined analogously to the continuous counterparts. Furthermore, denote by $\pi_h : L_2(\Omega)^2 \to V_h$ the L_2-orthogonal projection onto V_h. The discretized, reduced optimal control problem reads

$$\min_{\vec{f}_h \in V_h} \hat{J}_h(\vec{f}_h) = \frac{\gamma_u}{2} \|\mathcal{U}_h \vec{f}_h - \vec{u}_d^h\|^2 + \frac{\gamma_p}{2} \|\mathcal{P}_h \vec{f}_h - p_d^h\|^2 + \frac{\beta}{2} \|\vec{f}_h\|^2, \tag{2.4}$$

where $\vec{u}_d^h$, p_d^h are the L_2-projections of the data onto V_h, respectively M_h.

Let us investigate the structure of the algebraic system associated to the discretized optimal control problem. Let $\{\vec{\varphi}_j\}_{j=1}^p$ and $\{\psi_k\}_{k=1}^m$ be the basis functions of the spaces V_h and M_h, respectively. Furthermore, assume that $\{\vec{\varphi}_j\}_{j=1}^n$ are the basis functions of V_h^0 for some $n < p$. We expand the discrete solutions $\vec{u}_h$ and p_h as

$$\vec{u}_h(x) = \sum_{j=1}^n \mathbf{u}_j \vec{\varphi}_j(x), \quad p_h(x) = \sum_{k=1}^m \mathbf{p}_k \psi_k(x),$$

and we approximate the control by $\vec{f}_h(x) = \sum_{j=1}^{p} \mathbf{f}_j \vec{\varphi}_j(x)$. Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{B} \in \mathbb{R}^{m\times n}$ be the matrices related to the bilinear forms, $\mathbf{A} = [a_{ij}] = a(\vec{\varphi}_i, \vec{\varphi}_j)$, $\mathbf{B} = [b_{ij}] = b(\vec{\varphi}_i, \psi_j)$. Furthermore consider the mass-matrix $\mathbf{M}_{\vec{f}} = [\int_\Omega \vec{\varphi}_i \cdot \vec{\varphi}_j] \in \mathbb{R}^{p\times p}$ and its submatrices $\mathbf{M}_{\vec{u}\vec{f}} = \mathbf{M}_{\vec{f}}(1:n, 1:p) \in \mathbb{R}^{n\times p}$ and $\mathbf{M}_{\vec{u}} = \mathbf{M}_{\vec{f}}(1:n, 1:n) \in \mathbb{R}^{n\times n}$. The discrete state equation takes the form

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{M}_{\vec{u}\vec{f}} \\ 0 \end{bmatrix} \mathbf{f}.$$

After discretizing, the reduced cost functional becomes

$$\hat{J}_h = \frac{\gamma_u}{2}(\mathbf{u} - \mathbf{u}_d)^T \mathbf{M}_{\vec{u}}(\mathbf{u} - \mathbf{u}_d) + \frac{\gamma_p}{2}(\mathbf{p} - \mathbf{p}_d)^T \mathbf{M}_p(\mathbf{p} - \mathbf{p}_d) + \frac{\beta}{2}\mathbf{f}^T \mathbf{M}_{\vec{f}}\mathbf{f},$$

where $\mathbf{M}_p = [\int_\Omega \psi_i \psi_j]$ and $\mathbf{u}_d$, $\mathbf{p}_d$ are the coefficient vectors in the expansions of $\vec{u}_h^d$ and p_h^d in V_h^0, respectively M_h.

Let us introduce the matrices

$$\mathbf{S} = \begin{bmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & 0 \end{bmatrix}, \ \mathbf{M} = \begin{bmatrix} \gamma_u \mathbf{M}_{\vec{u}} & 0 \\ 0 & \gamma_p \mathbf{M}_p \end{bmatrix}, \ \mathbf{L} = \begin{bmatrix} \mathbf{M}_{\vec{u}\vec{f}} \\ 0 \end{bmatrix}, \text{ and } \quad \mathbf{x}_d = \begin{bmatrix} \mathbf{u}_d \\ \mathbf{p}_d \end{bmatrix}.$$

After dropping the constant terms the problem becomes

$$\min_{\mathbf{f}} \frac{1}{2}\mathbf{f}^T \big(\mathbf{L}^T \mathbf{S}^{-1} \mathbf{M} \mathbf{S}^{-1} \mathbf{L} + \beta \mathbf{M}_{\vec{f}}\big)\mathbf{f} - \mathbf{f}^T \mathbf{L}^T \mathbf{S}^{-1} \mathbf{M}\, \mathbf{x}_d,$$

which reduces to solving the linear system

$$\big(\mathbf{L}^T \mathbf{S}^{-1} \mathbf{M} \mathbf{S}^{-1} \mathbf{L} + \beta \mathbf{M}_{\vec{f}}\big)\mathbf{f} = \mathbf{L}^T \mathbf{S}^{-1} \mathbf{M}\, \mathbf{x}_d. \tag{2.5}$$

Note that the system matrix is dense, thus (2.5) has to be solved using iterative methods, and for increased efficiency we need high-quality preconditioners. In Section 3, we construct and analyze a multigrid preconditioner for the system

$$\big(\beta \mathbf{I} + \mathbf{M}_{\vec{f}}^{-1} \mathbf{L}^T \mathbf{S}^{-1} \mathbf{M} \mathbf{S}^{-1} \mathbf{L}\big)\mathbf{f} = \mathbf{M}_{\vec{f}}^{-1} \mathbf{L}^T \mathbf{S}^{-1} \mathbf{M}\, \mathbf{x}_d \ , \tag{2.6}$$

which is obtained from (2.5) by left-multiplying with $\mathbf{M}_{\vec{f}}^{-1}$.

We should remark that the system (2.6) can be obtained from the KKT system associated with the discrete constrained optimal control problem associated to (1.1)-(1.2) by block-eliminating the velocity, pressure, and the Lagrange multipliers. So (2.6) is in fact a reduced KKT system.

2.2. Estimates and convergence results. We first list some results on the Stokes equations and their numerical approximation which are needed for the multigrid analysis.

THEOREM 2.2. *There exist constants $C_1 = C_1(\mathcal{U}, \Omega)$ and $C_2 = C_2(\mathcal{P}, \Omega)$ such that the following hold:*

(a) smoothing:

$$\|\mathcal{U}\vec{f}\| \le C_1 \|\vec{f}\|_{H^{-2}(\Omega)^2} \quad \forall \vec{f} \in L_2(\Omega)^2, \tag{2.7}$$

and

$$\|\mathcal{P}\vec{f}\| \le C_2 \|\vec{f}\|_{H^{-1}(\Omega)^2} \quad \forall \vec{f} \in L_2(\Omega)^2; \tag{2.8}$$

(b) approximation:

$$\|\mathcal{U}\vec{f} - \mathcal{U}_h\vec{f}\| \le C_1 h^2 \|\vec{f}\| \quad \forall \vec{f} \in L_2(\Omega)^2, \tag{2.9}$$

and

$$\|\mathcal{P}\vec{f} - \mathcal{P}_h\vec{f}\| \le C_2 h \|\vec{f}\| \quad \forall \vec{f} \in L_2(\Omega)^2; \tag{2.10}$$

(c) stability:

$$\|\mathcal{U}_h\vec{f}\| \le C_1\|\vec{f}\| \quad \textit{and} \quad \|\mathcal{P}_h\vec{f}\| \le C_2\|\vec{f}\| \quad \forall \vec{f} \in L_2(\Omega)^2. \tag{2.11}$$

Proof.

(a) The inequality (2.7) is a straightforward consequence of (2.1) (see [5, Corollary 6.2]), while (2.8) follows immediately from Brezzi's splitting theorem, see, e.g., [2, Theorem 4.3].

(b) This is a standard approximation result, see, e.g., [13].

(c) Follows immediately from the estimates in (a) and (b).

□

We state without proof the following well-known result.

Lemma 2.3. *The following approximation of the identity by the projection holds:*

$$\|(I - \pi_h)\vec{f}\|_{H^{-k}(\Omega)^2} \le Ch^k\|\vec{f}\| \quad \forall \vec{f} \in L_2(\Omega)^2, \tag{2.12}$$

for $k = 0, 1, 2, 3$, with C constant independent of h.

Remark 2.4. *Theorem 2.2 and Lemma 2.3 imply that there are constants $C_1 = C_1(\Omega, \mathcal{U})$ and $C_2 = C_2(\Omega, \mathcal{P})$ such that*

$$\|\mathcal{U}(I - \pi_h)\vec{f}\| \le C_1 h^2 \|\vec{f}\| \quad \forall \vec{f} \in L_2(\Omega)^2 \tag{2.13}$$

and

$$\|\mathcal{P}(I - \pi_h)\vec{f}\| \le C_2 h \|\vec{f}\| \quad \forall \vec{f} \in L_2(\Omega)^2. \tag{2.14}$$

Lemma 2.5. *Let $\tilde{\pi}_h : L_{2,0}(\Omega) \to M_h$ be the L_2-orthogonal projection onto M_h. Then*

$$|((I - \tilde{\pi}_h)q, \mathcal{P}\vec{f})| \le Ch\|q\|\,\|\vec{f}\| \quad \forall q \in L_{2,0}(\Omega), \vec{f} \in L_2(\Omega)^2, \tag{2.15}$$

with C constant independent of h. *Proof.* We have

$$\begin{aligned} |((I - \tilde{\pi}_h)q, \mathcal{P}\vec{f})| &= |((I - \tilde{\pi}_h)q, \mathcal{P}\vec{f} - \mathcal{P}_h\vec{f})| \le \|(I - \tilde{\pi}_h)q\|\,\|\mathcal{P}\vec{f} - \mathcal{P}_h\vec{f}\| \\ &\overset{(2.10)}{\le} Ch\|q\|\,\|\vec{f}\| \quad \forall q \in L_{2,0}(\Omega), \vec{f} \in L_2(\Omega)^2. \qquad \square \end{aligned}$$

Denote

$$G_u = \mathcal{U}^*\mathcal{U}\ , \quad G_p = \mathcal{P}^*\mathcal{P}\ , \quad G_u^h = \mathcal{U}_h^*\mathcal{U}_h\ , \quad G_p^h = \mathcal{P}_h^*\mathcal{P}_h\ .$$

Lemma 2.6. *The following approximation properties hold:*

$$\|\pi_h(G_u^h - G_u)\vec{f}\| \le Ch^2\|\vec{f}\| \quad \forall \vec{f} \in V_h, \tag{2.16}$$

and

$$\|\pi_h(G_p^h - G_p)\vec{f}\| \le Ch\|\vec{f}\| \quad \forall \vec{f} \in V_h, \tag{2.17}$$

for some constant C independent of h.

Proof. For $\vec{f} \in V_h$ we have

$$\begin{aligned} |(\pi_h(G_u^h - G_u)\vec{f}, \vec{f})| &= |\|\mathcal{U}_h\vec{f}\|^2 - \|\mathcal{U}\vec{f}\|^2| \\ &\le \|\mathcal{U}_h\vec{f} - \mathcal{U}\vec{f}\|(\|\mathcal{U}_h\vec{f}\| + \|\mathcal{U}\vec{f}\|) \\ &\le Ch^2\|\vec{f}\|^2, \end{aligned}$$

and (2.16) follows from the symmetry of $\pi_h(G_u^h - G_u)$.

Similarly, we have

$$\begin{aligned} |(\pi_h(G_p^h - G_p)\vec{f}, \vec{f})| &= |(G_p^h\vec{f}, \vec{f})_V - (G_p\vec{f}, \vec{f})_V| \\ &= |(\mathcal{P}_h\vec{f}, \mathcal{P}_h\vec{f})_M - (\mathcal{P}\vec{f}, \mathcal{P}\vec{f})_M| \\ &= |\|\mathcal{P}_h\vec{f}\|^2 - \|\mathcal{P}\vec{f}\|^2| \\ &\le \|\mathcal{P}_h\vec{f} - \mathcal{P}\vec{f}\|(\|\mathcal{P}_h\vec{f}\| + \|\mathcal{P}\vec{f}\|) \\ &\le Ch\|\vec{f}\|^2 \quad \forall \vec{f} \in V_h, \end{aligned}$$

and (2.17) follows from the symmetry of $\pi_h(G_p^h - G_p)$. □

From the definition of H_β follows

$$\beta\|\vec{f}\|^2 \le (H_\beta\vec{f}, \vec{f}) \le (\beta + C)\|\vec{f}\|^2 \quad \forall \vec{f} \in L_2(\Omega)^2, \tag{2.18}$$

with $C = C(\mathcal{U}, \mathcal{P}, \Omega)$, and a similar estimate holds for the discrete Hessian

$$H_\beta^h = \beta I + \gamma_u G_u^h + \gamma_p G_p^h, \tag{2.19}$$

which shows that the condition number of H_β^h is bounded uniformly with respect to h, but potentially increasing with $\beta \downarrow 0$. The inequality (2.18) also implies that the cost functional $\hat{J}$ is strictly convex and has a unique minimizer given by

$$\vec{f}^{min} = H_\beta^{-1}(\gamma_u\mathcal{U}^*\vec{u}_d + \gamma_p\mathcal{P}^*p_d).$$

Similarly, the minimizer of the discrete quadratic is

$$\vec{f}_h^{min} = (H_\beta^h)^{-1}(\gamma_u\mathcal{U}_h^*\vec{u}_d^h + \gamma_p\mathcal{P}_h^*p_d^h).$$

In the following theorem, we show that $\vec{f}_h^{min}$ approximates $\vec{f}^{min}$ to optimal order in the L_2-norm.

THEOREM 2.7. *There exists a constant $C = C(\Omega, \mathcal{U}, \mathcal{P})$ independent of h such that for $h \le h_0(\beta, \Omega, \mathcal{U}, \mathcal{P})$ we have the following stability and error estimates:*

$$\|\vec{f}_h^{\min}\| \le \|\vec{f}^{\min}\| + \frac{C}{\beta}\big(\gamma_u h^2\|\vec{u}_d\| + \gamma_p h\|p_d\|\big), \tag{2.20}$$

$$\|\vec{f}_h^{\min} - \vec{f}^{\min}\| \le \frac{C}{\beta}\big(\gamma_u h^2(\|\vec{u}_d\| + \|\vec{f}^{min}\|) + \gamma_p h(\|p_d\| + \|\vec{f}^{min}\|)\big). \tag{2.21}$$

Proof. Let $\vec{e}_h = \vec{f}_h^{min} - \vec{f}^{min}$. We have

$$\begin{aligned}
H_\beta \vec{e}_h &= (\beta I + \gamma_u G_u + \gamma_p G_p)\vec{f}_h^{min} - H_\beta \vec{f}^{min}\\
&= (\beta I + \gamma_u G_u + \gamma_p G_p)\vec{f}_h^{min} - \gamma_u \mathcal{U}^* \vec{u}_d - \gamma_p \mathcal{P}^* p_d\\
&= (\beta I + \gamma_u G_u^h + \gamma_p G_p^h)\vec{f}_h^{min} + \gamma_u (G_u - G_u^h)\vec{f}_h^{min} + \gamma_p (G_p - G_p^h)\vec{f}_h^{min}\\
&\quad - \gamma_u \mathcal{U}^* \vec{u}_d - \gamma_p \mathcal{P}^* p_d\\
&= \gamma_u \mathcal{U}_h^* \vec{u}_d^h + \gamma_p \mathcal{P}_h^* p_d^h + \gamma_u (G_u - G_u^h)\vec{f}_h^{min} + \gamma_p (G_p - G_p^h)\vec{f}_h^{min}\\
&\quad - \gamma_u \mathcal{U}^* \vec{u}_d - \gamma_p \mathcal{P}^* p_d\\
&= \gamma_u \underbrace{\left[(G_u - G_u^h)\vec{f}_h^{min} + (\mathcal{U}_h^* - \mathcal{U}^*)\vec{u}_d^h - \mathcal{U}^*(I - \pi_h)\vec{u}_d\right]}_{\mathcal{A}_u}\\
&\quad + \gamma_p \underbrace{\left[(G_p - G_p^h)\vec{f}_h^{min} + (\mathcal{P}_h^* - \mathcal{P}^*)p_d^h - \mathcal{P}^*(I - \tilde{\pi}_h)p_d\right]}_{\mathcal{A}_p},
\end{aligned}$$

where we used that $\vec{u}_d^h = \pi_h \vec{u}_d$ and $p_d^h = \tilde{\pi}_h p_d$. Furthermore,

$$\beta \|\vec{e}_h\|^2 \overset{(2.18)}{\leq} (H_\beta \vec{e}_h, \vec{e}_h) \leq \gamma_u (\mathcal{A}_u, \vec{e}_h) + \gamma_p (\mathcal{A}_p, \vec{e}_h),$$

and

$$\begin{aligned}
(\mathcal{A}_u, \vec{e}_h) &\overset{(2.9),(2.13)}{\leq} Ch^2 \|\vec{u}_d\| \|\vec{e}_h\| + ((G_u - G_u^h)\vec{f}_h^{min}, \vec{e}_h)\\
&\overset{(I-\pi_h)e_h \perp V_h}{=} Ch^2 \|\vec{u}_d\| \|\vec{e}_h\| + (\pi_h (G_u - G_u^h)\vec{f}_h^{min}, \pi_h \vec{e}_h)\\
&\quad + (G_u \vec{f}_h^{min}, (I - \pi_h)\vec{e}_h)\\
&\overset{(2.16)}{\leq} Ch^2 \|\vec{e}_h\| (\|\vec{u}_d\| + \|\vec{f}_h^{min}\|) + (\mathcal{U} \vec{f}_h^{min}, \mathcal{U}(I - \pi_h)\vec{e}_h)\\
&\overset{(2.13)}{\leq} Ch^2 \|\vec{e}_h\| (\|\vec{u}_d\| + \|\vec{f}_h^{min}\|),
\end{aligned}$$

with $C = C(\mathcal{U}, \Omega)$, where we have also used the fact that $\mathcal{U}$ is self-adjoint. Similarly, we get

$$\begin{aligned}
(\mathcal{A}_p, \vec{e}_h) &= (p_d^h, (\mathcal{P}_h - \mathcal{P})\vec{e}_h) - ((I - \tilde{\pi}_h)p_d, \mathcal{P}\vec{e}_h) + ((G_p - G_p^h)\vec{f}_h^{min}, \vec{e}_h)\\
&\overset{(2.10),(2.15)}{\leq} Ch \|p_d\| \|\vec{e}_h\| + ((G_p - G_p^h)\vec{f}_h^{min}, \vec{e}_h)\\
&\overset{(I-\pi_h)\vec{e}_h \perp V_h}{=} Ch \|p_d\| \|\vec{e}_h\| + (\pi_h (G_p - G_p^h)\vec{f}_h^{min}, \pi_h \vec{e}_h)\\
&\quad + (G_p \vec{f}_h^{min}, (I - \pi_h)\vec{e}_h)\\
&\overset{(2.17)}{\leq} Ch \|\vec{e}_h\| (\|p_d\| + \|\vec{f}_h^{min}\|) + (\mathcal{P} \vec{f}_h^{min}, \mathcal{P}(I - \pi_h)\vec{e}_h)\\
&\overset{(2.14)}{\leq} Ch \|\vec{e}_h\| (\|p_d\| + \|\vec{f}_h^{min}\|),
\end{aligned}$$

with $C = C(\mathcal{P}, \Omega)$. These estimates yield

$$\|\vec{f}_h^{min} - \vec{f}^{min}\| \leq \frac{C}{\beta}\Big(\gamma_u h^2 (\|u_d\| + \|\vec{f}_h^{min}\|) + \gamma_p h (\|p_d\| + \|\vec{f}_h^{min}\|)\Big).$$

Since

$$\begin{aligned}\|\vec{f}_h^{min}\| &\leq \|\vec{f}^{min}\| + \|\vec{f}_h^{min} - \vec{f}^{min}\| \\ &\leq \|\vec{f}^{min}\| + \frac{C}{\beta}\Big(\gamma_u h^2(\|u_d\| + \|\vec{f}_h^{min}\|) + \gamma_p h(\|p_d\| + \|\vec{f}_h^{min}\|\Big),\end{aligned}$$

we obtain (2.20) and (2.21) for h sufficiently small. □

3. Two-grid and multigrid preconditioner for the discrete Hessian. In this section, we use the multigrid techniques developed in [5] to construct and analyze a two-level symmetric preconditioner for the discrete Hessian H_β^h defined in (2.19). The extension to a multigrid preconditioner follows the same strategy as in [5] and is further explained in great detail in [6].

For the remainder of this paper we consider on V_h the Hilbert-space structure inherited from $L_2(\Omega)^2$. Furthermore, we consider the L_2-orthogonal decomposition $V_h = V_{2h} \oplus W_{2h}$ and let π_{2h} be the L_2-projector onto V_{2h}. The analysis in [5] suggests that H_β^h is well approximated by

$$T_\beta^h \stackrel{\text{def}}{=} H_\beta^{2h}\pi_{2h} + \beta(I - \pi_{2h}).$$

For completeness we briefly recall the heuristics leading to the definition of T_β^h. As usual in the multigrid literature, for $\vec{f}_h \in V_h$ we regard $\pi_{2h}\vec{f}_h$ as the "smooth" component of $\vec{f}_h$, and $(I - \pi_{2h})\vec{f}_h$ as the "rough" or "oscillatory" component; so the projector $(I - \pi_{2h})$ extracts the "oscillatory" part of a function in V_h. If we write $H_\beta^h = H_0^h + \beta I$ and take into account the "smoothing" properties of H_0^h (these are due to the compactness of the operator H_0 which H_0^h approximates), it follows that the products $(I - \pi_{2h})H_0^h$ and $H_0^h(I - \pi_{2h})$ are almost negligible. So

$$\begin{aligned}H_\beta^h &= (\pi_{2h} + (I - \pi_{2h}))(H_0^h + \beta I)(\pi_{2h} + (I - \pi_{2h})) \\ &\approx \pi_{2h}(H_0^h + \beta I)\pi_{2h} + \beta(I - \pi_{2h}) \ . \qquad (3.1)\end{aligned}$$

Furthermore, when applied to the "smooth" component $\pi_{2h}\vec{f}_h$ of a function $\vec{f}_h$, it is expected that $H_0^h\pi_{2h}\vec{f}_h \approx H_0\pi_{2h}\vec{f}_h \approx H_0^{2h}\pi_{2h}\vec{f}_h$, hence the idea to replace in (3.1) H_0^h by H_0^{2h}, which gives rise to T_β^h.

Since π_{2h} is a projection, $(T_\beta^h)^{-1}$ is computed explicitly as

$$L_\beta^h \stackrel{\text{def}}{=} (T_\beta^h)^{-1} = (H_\beta^{2h})^{-1}\pi_{2h} + \beta^{-1}(I - \pi_{2h}).$$

We propose $L_\beta^h \in \mathcal{L}(V_h)$ as a two-level preconditioner for H_β^h. To assess the quality of the preconditioner we use the spectral distance introduced in [5], defined for two symmetric positive definite operators $T_1, T_2 \in \mathcal{L}(V_h)$ as

$$d_h(T_1, T_2) = \sup_{w \in V_h \setminus \{0\}} \left| \ln \frac{(T_1 w, w)}{(T_2 w, w)} \right|. \qquad (3.2)$$

If L_β^h is a preconditioner for H_β^h then the spectral radius $\rho(I - L_\beta^h H_\beta^h)$, which is an accepted quality-measure for a preconditioner, is controlled by the spectral distance between L_β^h and $(H_\beta^h)^{-1}$ (see Lemma A.2 in Appendix A for a precise formulation). The advantage of using the spectral distance over $\rho(I - L_\beta^h H_\beta^h)$ is that the former is a true distance function.

Theorem 3.1. *For $h < h_0(\beta,\Omega,\mathcal{U},\mathcal{P})$ there exists a constant $C = C(\Omega,\mathcal{U},\mathcal{P})$ such that*

$$d_h(H_\beta^h, T_\beta^h) \leq \frac{C}{\beta}(\gamma_u h^2 + \gamma_p h). \tag{3.3}$$

Proof. We have

$$\begin{aligned} T_\beta^h - H_\beta^h &= H_\beta^{2h}\pi_{2h} + \beta(I - \pi_{2h}) - H_\beta^h \\ &= (\beta I + \gamma_u G_u^{2h} + \gamma_p G_p^{2h})\pi_{2h} + \beta(I - \pi_{2h}) - (\beta I + \gamma_u G_u^h + \gamma_p G_p^h) \\ &= \gamma_u(G_u^{2h}\pi_{2h} - G_u^h) + \gamma_p(G_p^{2h}\pi_{2h} - G_p^h). \end{aligned} \tag{3.4}$$

We write

$$G_u^{2h}\pi_{2h} - G_u^h = (G_u^{2h} - G_u)\pi_{2h} + (G_u - G_u^h)\pi_{2h} + G_u^h(\pi_{2h} - I).$$

For any $\vec{g} \in V_h$ we have

$$\begin{aligned} |((G_u^{2h} - G_u)\vec{g}, \vec{g})| &= |(\mathcal{U}_{2h}^*\mathcal{U}_{2h} - \mathcal{U}^*\mathcal{U})\vec{g}, \vec{g})| = |\|U_{2h}\vec{g}\|^2 - \|\mathcal{U}\vec{g}\|^2| \\ &\leq \|(\mathcal{U}_{2h} - \mathcal{U})\vec{g}\| \; (\|\mathcal{U}_{2h}\vec{g}\| + \|\mathcal{U}\vec{g}\|) \overset{(2.9)}{\leq} Ch^2\|\vec{g}\|^2, \end{aligned}$$

which implies

$$\|(G_u^{2h} - G_u)\vec{g}\| \leq Ch^2\|\vec{g}\|$$

since $G_u^{2h} - G_u$ is symmetric on V_h. In particular,

$$\|(G_u^{2h} - G_u)\pi_{2h}\vec{f}\| \leq Ch^2\|\vec{f}\| \quad \forall \vec{f} \in V_h.$$

Similarly, it can be shown that

$$\|(G_u - G_u^h)\pi_{2h}\vec{f}\| \leq Ch^2\|\vec{f}\|.$$

Finally, we estimate

$$\begin{aligned} \|G_u^h(I - \pi_{2h})\vec{f}\| &= \|\mathcal{U}_h^*\mathcal{U}_h(I - \pi_{2h})\vec{f}\| \leq C\|\mathcal{U}_h(I - \pi_{2h})\vec{f}\| \\ &\leq C\|\mathcal{U}(I - \pi_{2h})\vec{f}\| + C\|(\mathcal{U} - \mathcal{U}_h)(I - \pi_{2h})\vec{f}\| \\ &\overset{(2.13),(2.9)}{\leq} C_1h^2\|\vec{f}\| + C_2h^2\|(I - \pi_{2h})\vec{f}\| \leq Ch^2\|\vec{f}\|, \quad \forall \vec{f} \in V_h. \end{aligned}$$

Combining the last three estimates, we obtain

$$\|(G_u^{2h}\pi_{2h} - G_u)\vec{f}\| \leq Ch^2\|\vec{f}\| \quad \forall \vec{f} \in V_h.$$

The second term in (3.4) is estimated in a similar way, to obtain

$$\|(G_p^{2h}\pi_{2h} - G_p)\vec{f}\| \leq Ch\|\vec{f}\| \quad \forall \vec{f} \in V_h,$$

which together with the previous estimate yields

$$\|(T_\beta^h - H_\beta^h)\vec{f}\| \leq C(\gamma_u h^2 + \gamma_p h)\|\vec{f}\| \quad \forall \vec{f} \in V_h.$$

It follows that

$$\left|\frac{(T_\beta^h \vec{f},\vec{f})}{(H_\beta^h \vec{f},\vec{f})}-1\right| \le C\frac{(\gamma_u h^2+\gamma_p h)\|\vec{f}\|^2}{\gamma_u(H_u^h\vec{f},\vec{f})+\gamma_p(H_p^h\vec{f},\vec{f})+\beta\|\vec{f}\|^2} \le \frac{C}{\beta}(\gamma_u h^2+\gamma_p h).$$

Assuming $C\beta^{-1}(\gamma_u h_0^2+\gamma_p h_0)=\alpha<1$, and $0<h\le h_0$ we conclude that T_β^h is symmetric positive definite, and

$$\begin{aligned}\sup_{\vec{f}\in V_h\setminus\{0\}}\left|\ln\frac{(T_\beta^h \vec{f},\vec{f})}{(H_\beta^h \vec{f},\vec{f})}\right| &\le \frac{|\ln(1-\alpha)|}{\alpha}\sup_{\vec{f}\in V_h\setminus\{0\}}\left|\frac{(T_\beta^h \vec{f},\vec{f})}{(H_\beta^h \vec{f},\vec{f})}-1\right|\\ &\le \frac{|\ln(1-\alpha)|}{\alpha}\frac{C}{\beta}(\gamma_u h^2+\gamma_p h)\\ &\le \frac{C}{\beta}(\gamma_u h^2+\gamma_p h),\end{aligned}$$

where we also used that for $\alpha\in(0,1)$, $x\in[1-\alpha,1+\alpha]$ we have

$$\frac{\ln(1+\alpha)}{\alpha}|1-x|\le|\ln x|\le\frac{|\ln(1-\alpha)|}{\alpha}|1-x|. \qquad \square$$

A consequence of Theorem 3.1, that legitimizes the use of L_β^h as a preconditioner for the Hessian, is stated in the following corollary.

Corollary 3.2. *There exists a constant $C=C(\Omega,\mathcal{U},\mathcal{P})$ such that*

$$d_h(L_\beta^h,(H_\beta^h)^{-1})\le\frac{C}{\beta}(\gamma_u h^2+\gamma_p h). \tag{3.5}$$

for $h\le h_0(\beta,\Omega,\mathcal{U},\mathcal{P})$.

Note that, for fixed β, since the spectral distance between the operators decreases with $h\downarrow 0$, the quality of the preconditioner increases with $h\downarrow 0$; this is different from classical multigrid preconditioners for elliptic problems, where the spectral distance is bounded above by a constant that is independent of h.

Next, we briefly describe how to extend the two-grid preconditioner to a multigrid preconditioner that exhibits the same optimal-order quality (3.5) and is less costly to apply. If we use the classical V-cycle idea to define recursively a multigrid preconditioner, the resulting preconditioner is suboptimal, that is, the quality of the preconditioner does not improve as $h\downarrow 0$, it is simply mesh-independent. To construct an improved preconditioner we introduce the operators

$$\mathcal{G}_h:\mathcal{L}(V_{2h})\to\mathcal{L}(V_h),\quad \mathcal{G}_h(T)=T\pi_{2h}+\beta^{-1}(I-\pi_{2h})$$

and

$$\mathcal{N}_h:\mathcal{L}(V_h)\to\mathcal{L}(V_h),\quad \mathcal{N}_h(X)=2X-XH_\beta^hX.$$

Note that $\mathcal{N}_h$ is related to the Newton iterator for the equation $X^{-1}-H_\beta^h=0$, i.e., $\mathcal{N}_h(X_0)$ is the first Newton iterate starting at X_0. Thus, if X_0 is a good approximation for $(H_\beta^h)^{-1}$ then $\mathcal{N}_h(X_0)$ is significantly closer to $(H_\beta^h)^{-1}$ than X_0. We follow [6] and define the multigrid preconditioner K_β^h using the following algorithm:

```
1. if coarsest level
                         $K^h_\beta = (H^h_\beta)^{-1}$
2. else
     if intermediate level
                         $K^h_\beta = \mathcal{N}_h(\mathcal{G}_h(K^{2h}_\beta))$
     else % finest level
                         $K^h_\beta = \mathcal{G}_h(K^{2h}_\beta)$
     end if
3. end if
```

Since the application of $\mathcal{N}_h$ requires a matrix-vector multiplication by H^h_β, which for large scale problems is expected to be very costly at the finest level, we prefer that no such matrix-vector multiplication is computed inside the preconditioner. This is the reason why we treat the cases of intermediate and finest resolutions differently. In practice, neither H^h_β nor K^h_β is ever formed, so both are applied matrix-free (see [6] for details).

The analysis of the multigrid preconditioner relies on the estimate for the two-grid preconditioner and properties of the spectral distance. We recall here Lemma 5.3 from [5] that is needed for the analysis.

Lemma 3.3. *Let $(e_i)_{i\geq 0}$ and $(a_i)_{i\geq 0}$ be positive numbers satisfying the recursive inequality*

$$e_{i+1} \leq C(e_i + a_{i+1})^2$$

and

$$a_{i+1} \leq a_i \leq f^{-1} a_{i+1}$$

for some $0 < f < 1$. If $a_0 \leq \frac{f}{4C}$ and if $e_0 \leq 4Ca_0^2$, then

$$e_i \leq 4Ca_i^2, \quad \forall i > 0.$$

Theorem 3.4. *Assume that $C\beta^{-1}(\gamma_u h_0^2 + \gamma_p h_0) < 2^{-5}$, where C is the constant from Theorem 3.1 and $K^{h_0}_\beta = (H^{h_0}_\beta)^{-1}$. Then*

$$d_h\big(K^h_\beta, (H^h_\beta)^{-1}\big) \leq 2\frac{C}{\beta}(\gamma_u h^2 + \gamma_p h), \quad \text{for } h = 2^{-l}h_0,\ l \geq 2. \tag{3.6}$$

Proof. Let $h_i = 2^{-i}h_0$, $i = 0, \dots, l$, and $e_i = e_{h_i} = d_{h_i}(K^{h_i}_\beta, (H^{h_i}_\beta)^{-1})$, $a_i = C\beta^{-1}(\gamma_u h_i + \gamma_p h_i)$. We have

$$\begin{aligned} d_{h_i}\big(\mathcal{G}(K^{2h_i}_\beta), (H^{h_i}_\beta)^{-1}\big) &\leq d_{h_i}\big(\mathcal{G}(K^{2h_i}_\beta), \mathcal{G}((H^{2h_i}_\beta)^{-1})\big) \\ &\quad + d_{h_i}\big(\mathcal{G}((H^{2h_i}_\beta)^{-1}), (H^{h_i}_\beta)^{-1}\big) \\ &\overset{(3.5)}{\leq} d_{h_i}\big(\mathcal{G}(K^{2h_i}_\beta), \mathcal{G}((H^{2h_i}_\beta)^{-1})\big) + \frac{C}{\beta}(\gamma_u h_i^2 + \gamma_p h_i) \qquad (3.7) \\ &\leq d_{2h_i}\big(K^{2h_i}_\beta, (H^{2h_i}_\beta)^{-1}\big) + \frac{C}{\beta}(\gamma_u h_i^2 + \gamma_p h_i) \\ &\leq e_{i-1} + a_i, \quad i = 1, \dots, l, \end{aligned}$$

where we have used the fact that $d_{h_i}(\mathcal{G}_{h_i}(T_1), \mathcal{G}_{h_i}(T_2)) \leq d_{2h_i}(T_1, T_2)$, for any two symmetric positive definitive operators $T_1, T_2 \in \mathcal{L}(V_{2h_i})$ [5, Lemma 5.1].

It is shown in [5] that for any $M, H \in \mathcal{L}_+(V_{h_i})$, with $d_{h_i}(M, H^{-1}) < 0.4$, we have

$$d_{h_i}(\mathcal{N}_{h_i}(M), H^{-1}) \leq 2 d_{h_i}(M, H^{-1})^2, \tag{3.8}$$

for $h_i < h_0$. An inductive argument implies that $e_i \leq 0.2$ for all i, provided it holds for e_0 and $C\beta^{-1}(\gamma_u h_0^2 + \gamma_p h_0) < 0.1$. Thus, combining (3.8) and (3.7) we obtain

$$e_i \leq 2(e_{i-1} + a_i)^2, \quad i = 1, \dots l - 1.$$

Note that the hypothesis of Lemma 3.1 are satisfied with $f = 1/4$ which implies

$$e_i \leq 8 \frac{C^2}{\beta^2} (\gamma_u h_i^2 + \gamma_p h_i)^2, \quad i = 1, \dots, l - 1.$$

In particular, we have

$$e_{2h} \leq 32 \frac{C^2}{\beta^2} (\gamma_u h^2 + \gamma_p h)^2.$$

At the finest level, $i = l$, (3.7) becomes

$$d_h\big(K_\beta^h, (H_\beta^h)^{-1}\big) \leq e_{2h} + \frac{C}{\beta}(\gamma_u h^2 + \gamma_p h),$$

which combined with the above estimate yields the assertion of the theorem. □

Finally, let us note that an important difference between our method and the classical multigrid is that here the base case needs to be chosen sufficiently fine, whereas in classical multigrid it can be as coarse as possible, in general.

4. Numerical results. In this section we present some numerical experiments that illustrate the application of the preconditioner introduced in Section 3.

Let $\Omega = (0, 1)^2$ and consider an optimal control problem of the form (1.1). We consider a family of uniform rectangular grids with mesh size h and discretize the problem using Taylor-Hood $\mathbf{Q}_2 - \mathbf{Q}_1$ elements for velocity-pressure and $\mathbf{Q}_2$ elements for the control. The problem was solved using MATLAB R2010A. We perform three types of experiments: velocity control only ($\gamma_u = 1, \gamma_p = 0$), mixed velocity-pressure control ($\gamma_u = 1, \gamma_p \neq 0$), and pressure control only ($\gamma_u = 0, \gamma_p = 1$).

First, we summarize the numerical results obtained for "in-vitro experiments". As mentioned earlier, the Hessian matrix is dense, therefore it is never formed in practice, and the matrix-vector products in the preconditioned conjugate gradient (PCG) are implemented matrix-free. However, in order to evaluate directly the spectral distance between the Hessian and the proposed two-level preconditioner, we formed the matrices for moderate values of the mesh size h. In Table 4.1 we present the joint spectrum analysis for $\beta = 1$, $\gamma_u = 1$, and $\gamma_p = 0$ with $d_h = \max\{|\ln \alpha| : \alpha \in \sigma(H_\beta^h, T_\beta^h)\}$, where $\sigma(A, B)$ denotes the set of generalized eigenvalues of A,B. The results indicate optimal third-order convergence. In Table 4.2 we present similar results for the case of pressure control only. In this case we observe an optimal quadratic convergence rate. We note that our computational results show a better behavior than predicted by Theorem 3.1. We think this is due to the particular type of convex domain chosen here, for which the Stokes problem has better regularity than what was assumed in Theorem 3.1, and also to the use of quadratic elements for the control.

Table 4.1

Joint spectrum analysis for $\beta = 1$: velocity control only ($\gamma_u = 1$, $\gamma_p = 0$).

h	d_h	d_{2h}/d_h
2^{-2}	1.0274×10^{-4}	N/A
2^{-3}	1.3308×10^{-5}	7.7205
2^{-4}	1.3883×10^{-6}	9.5858
2^{-5}	1.5834×10^{-7}	8.7675

We also remark that the numerical estimates of d_h in case of pressure control only were obtained with a discretization that represents faithfully the finite element formulation in Section 2.1. However, in practice, instead of using average-zero pressures it is convenient to use a pressure space where the pressures are set to zero at a fixed location, e.g., a corner. If we compute the operators H_β^h, T_β^h using the latter space, we note that $\sigma(H_\beta^h, T_\beta^h)$ contains exactly two generalized eigenvalues that are of size $O(1)$, instead of $O(h)$, as predicted by Theorem 3.1. If $\widetilde{\sigma}(H_\beta^h, T_\beta^h)$ is the subset of $\sigma(H_\beta^h, T_\beta^h)$ obtained after excluding the two generalized eigenvalues that are $O(1)$, and $\tilde{d}_h = \max\{|\ln\alpha| : \alpha \in \widetilde{\sigma}(H_\beta^h, T_\beta^h)\}$, then $\tilde{d}_h = O(h)$, as the theory predicts for d_h.

Table 4.2

Joint spectrum analysis for $\beta = 1$: pressure control only ($\gamma_u = 0$, $\gamma_p = 1$).

h	d_h	d_{2h}/d_h
2^{-2}	1.9613×10^{-2}	N/A
2^{-3}	5.6686×10^{-3}	3.4599
2^{-4}	1.6703×10^{-3}	3.3937
2^{-5}	4.3735×10^{-4}	3.8192

The next type of numerical experiments regard the solution of the control problem (1.1). Specifically, we compare the number of iterations required to solve the linear system (2.6) with unpreconditioned CG to the case when CG is used with a multilevel preconditioner with $1-4$ levels (depending on resolution, see more comments below). For the results presented here, we chose the target velocity

$$\vec{u}_d = (-2x^2y(1-x)^2(1-3y+2y^2), 2xy^2(1-y)^2(1-3x+2x^2))^T$$

and the target pressure

$$p_d = \cos\pi x \cos\pi y\ .$$

In Table 4.3 we summarize the results obtained for velocity control only ($\gamma_u = 1$, $\gamma_p = 0$), mixed velocity-pressure control ($\gamma_u = 1, \gamma_p = 10^{-5}, 10^{-4}, 10^{-3}$), and pressure control only ($\gamma_u = 0, \gamma_p = 1$); for each case a range of values for β is chosen.

The choice of the values for γ_p in the mixed velocity-pressure control is justified by the data in Table 4.4, where we show for each case (and fixed resolution $h = 1/32$) the relative error in the recovered data for the velocity $E_{\vec{u}} = \|\vec{u}_h^{\min} - \vec{u}_d\|/\|\vec{u}_d\|$ and pressure $E_p = \|p_h^{\min} - p_d\|/\|p_d\|$. As can be seen from Table 4.4 for pure velocity control, the pressure is not recovered at all ($E_p \approx 1$), while for pure pressure control the velocity is not recovered ($E_{\vec{u}} \approx 1$). As expected, for mixed control both pressure and velocity are being recovered in the sense that both $E_{\vec{u}}$ and E_p decrease with $\beta \downarrow 0$. However, for $\gamma_p = 10^{-5}$ the relative velocity error $E_{\vec{u}}$ is one order of magnitude smaller than E_p, so velocity is better recovered than pressure. For $\gamma_p = 10^{-4}$, $E_{\vec{u}}$ and E_p are of comparable size (within a factor of 2), and if $\gamma_p = 10^{-3}$, then the situation is reversed with E_p being one order of magnitude smaller than $E_{\vec{u}}$.

We now return to the actual results in Table 4.3. First note that for all cases unpreconditioned CG solves the system (2.6) in a number of iterations that appears to be almost mesh-independent, and is clearly bounded with $h \downarrow 0$. While this is certainly consistent with our remark from Section 2 regarding the mesh independence of the condition number of H_β^h, the relatively low number of CG iterations is due to the fact that the continuous counterpart of H_β^h is compact, which implies that the number of eigenvalues of H_β^h which are away from β is small. Also in accordance with (2.18), the number of unpreconditioned CG iterations increases with $\beta \downarrow 0$, which is expected since the number of relevant eigenvalues of H_β^h increases as $\beta \downarrow 0$. For velocity control only (the top part of Table 4.3) we observe a significant reduction in the number of iterations when multilevel preconditioners are used, as well as a decrease in the number of iterations with mesh size. For example, for $\beta = 10^{-7}$ and $h = 2^{-8}$, the number of preconditioned iterations with a four-level preconditioner is significantly smaller than in the unpreconditioned case, i.e, 3 iterations vs. 78 iterations. Although a preconditioned iteration is more expensive than an unpreconditioned one, for large problems the overall cost of the preconditioned solver is much lower than of the unpreconditioned one, as can be seen from Table 4.5. At the other end of these experiments (bottom of Table 4.3) we see the cases of pressure control only. We also see the decrease in number of iterations with $h \downarrow 0$ when comparing the behavior of, say, 3-grid preconditioners at different resolutions: for example, for $\gamma_u = 0, \gamma_p = 1, \beta = 10^{-3}$ the number of 3-level preconditioners dropped from 14 ($h = 2^{-6}$) to 10 ($h = 2^{-7}$); similar results are consistently observed throughout Table 4.3. However, for the case of pressure control only, the drop in number of iterations from unpreconditioned CG to multilevel preconditioned CG is not as dramatic as for velocity control only. As can be inferred from Table 4.3, the efficiency of the multilevel preconditioned CG versus unpreconditioned CG measured as the ratio of the number of iterations gradually decreases with the increase of the ratio γ_p/γ_u (for otherwise comparable experiments), as predicted by Theorems 3.1 and 3.4.

In our implementation we have used direct methods for solving the Stokes system, and we actually constructed the base-case Hessian and stored its inverse. These choices have limited our computations to $h = 2^{-8}$ and $h_{\text{base}} = 2^{-5}$, which is why for $h = 2^{-7}$ we were unable to test the two-grid preconditioner (this would have required saving a base-case Hessian for $h = 2^{-6}$), and for $h = 2^{-8}$ we were only able to compute using a four-level preconditioner. While we already commented on the positive side of the numerical results, it is worth noting the pitfalls: if the coarsest grid is too coarse, then the quality of the preconditioner declines to the point that it is hurting the computation. This can be seen in the groups of columns and rows in Table 4.3 corresponding to $h = 2^{-6}$ and $\beta = 10^{-6}, 10^{-7}$: the use of too many levels results in a spike in the number of iterations to the point of non-convergence (within a maximum number of 100 iterations allowed).

Finally, we would like to comment on the robustness of our algorithm with respect to the accuracy of the Stokes solve. For large-scale problems the Stokes system on the finer grids is expected to be solved using iterative rather than direct methods, which reduces the accuracy of computing matrix-vector multiplications for the Hessian matrix H^h_β. We have repeated our numerical experiments in that, except for the coarsest scale, we replaced the direct solve of the Stokes systems with a preconditioned MINRES solve, as described in [7]. For the case of pure velocity control ($\gamma_p = 0$) we found no significant change in the number of iterations in Table 4.3. However, for the case of mixed- and pure pressure control ($\gamma_p \neq 0$) the quality of our algorithm appeared to decline significantly. We identified as the primary cause for this behavior the fact that, even when the velocity variables are well resolved, that is, the relative error between the solutions obtained via direct vs. iterative methods is on the order of 10^{-8}, the relative error in the pressure terms can be quite high (10^{-2}–10^{-4}). Since our algorithm relies on the ability to compute the operator $\mathcal{P}_h$ with sufficient accuracy, we are not able at this point to draw conclusions with respect to the influence of using iterative methods on our algorithm for mixed- and pure pressure control.

Table 4.3
Iteration count for multilevel preconditioners; "nc" means "not-converged". The tolerance is set at 10^{-12}.

h	2^{-6}				2^{-7}			2^{-8}	
num. levels	1	2	3	4	1	3	4	1	4
$\gamma_u = 1, \quad \gamma_p = 0$									
$\beta = 10^{-4}$	7	3	3	3	7	2	2	7	2
$\beta = 10^{-5}$	13	3	3	3	13	3	3	14	2
$\beta = 10^{-6}$	29	4	4	6	29	3	3	32	3
$\beta = 10^{-7}$	74	5	7	62	75	3	5	78	3
$\gamma_u = 1, \quad \gamma_p = 10^{-5}$									
$\beta = 10^{-4}$	10	4	5	5	10	5	5	11	5
$\beta = 10^{-5}$	20	5	6	8	20	6	8	22	7
$\beta = 10^{-6}$	45	7	8	13	44	7	9	45	9
$\beta = 10^{-7}$	112	9	14	nc	113	9	14	119	11
$\gamma_u = 1, \quad \gamma_p = 10^{-4}$									
$\beta = 10^{-4}$	10	5	6	8	11	5	7	11	7
$\beta = 10^{-5}$	21	7	7	9	23	7	9	24	9
$\beta = 10^{-6}$	48	8	10	16	48	9	13	49	11
$\beta = 10^{-7}$	122	13	17	nc	126	11	22	133	20
$\gamma_u = 1, \quad \gamma_p = 10^{-3}$									
$\beta = 10^{-4}$	12	6	7	9	12	7	9	13	9
$\beta = 10^{-5}$	25	8	9	13	26	9	13	28	11
$\beta = 10^{-6}$	61	11	17	33	63	12	22	65	20
$\gamma_u = 0, \quad \gamma_p = 1$									
$\beta = 10^{-1}$	8	5	7	9	8	6	8	8	8
$\beta = 10^{-2}$	13	7	9	12	13	8	11	13	11
$\beta = 10^{-3}$	27	10	14	nc	26	10	15	27	14

5. Conclusions. In this article we introduced Schur-based two- and multigrid preconditioners for the KKT system associated to a Tikhonov-regularized optimal control problem constrained by the Stokes system. We showed that, if the Stokes-system is discretized using a stable pair of finite elements, the preconditioner approximates the reduced Hessian of KKT system to optimal order with respect to the convergence order of the finite element method. As a consequence, the number of preconditioned

Table 4.4
Relative error for recovered data for velocity control $E_{\vec{u}} = \|\vec{u}_h^{\min} - \vec{u}_d\|/\|\vec{u}_d\|$ and pressure $E_p = \|p_h^{\min} - p_d\|/\|p_d\|$.

γ_u	γ_p	$E_{\vec{u}}$	E_p	$E_{\vec{u}}$	E_p	$E_{\vec{u}}$	E_p
β		10^{-5}		10^{-6}		10^{-7}	
1	0	4.07e-2	≈ 1	8.23e-3	≈ 1	1.88e-3	≈ 1
1	10^{-5}	4.32e-2	4.32e-1	1.65e-2	2.80e-1	8.28e-3	7.05e-2
1	10^{-4}	6.66e-2	2.75e-1	3.30e-2	6.85e-2	9.55e-3	8.44e-3
1	10^{-3}	1.23e-1	6.56e-2	3.78e-2	8.18e-3	9.70e-3	8.98e-4
β		10^{-1}		10^{-2}		10^{-3}	
0	1	1.02	2.66e-1	1.07	6.12e-2	1.09	7.34e-3

Table 4.5
Velocity control only: time comparison for $h = 2^{-8}$. No. state variables: 588290, no. control variables: 522242.

no. levels	1	4
$\beta = 10^{-6}$	3613 s	493 s + 1486 s for base case
$\beta = 10^{-7}$	8953 s	575 s + 1492 s for base case

CG iterations needed for solving the optimal control problem to a given tolerance decreases with increasing resolution, asymptoting to just one iteration as $h \downarrow 0$. The problem discussed in this article forms an important stepping stone towards finding highly efficient methods for solving large-scale optimal control problems constrained by the Navier-Stokes system, which are the focus of our current research.

Appendix A. Some facts about spectral distance.

Let $(X, \langle\cdot,\cdot\rangle)$ be a real finite dimensional Hilbert space and denote the complexification of X by $X^{\mathbb{C}} = \{u + \mathbf{i}v : u, v \in X\}$. Let $\mathcal{L}_+(X) = \{T \in \mathcal{L}(X) : \langle Tu, u\rangle > 0,\ \forall u \in X \setminus \{0\}\}$ and define the spectral distance between $S, T \in \mathcal{L}_+(X)$ to be

$$d_X(S,T) = \sup_{w \in X^{\mathbb{C}} \setminus \{0\}} \left| \ln \frac{(S^{\mathbb{C}}w, w)}{(T^{\mathbb{C}}w, w)} \right|,$$

where $T^{\mathbb{C}}(u+\mathbf{i}v) = T(u)+\mathbf{i}T(v)$ is the complexification of T. The following inequalities were proved in [5, Lemma 3.2]:

Lemma A.1. *If $\alpha \in (0,1)$ and $z \in \mathcal{B}_\alpha(1)$, then*

$$\frac{\ln(1+\alpha)}{\alpha}|1-z| \le |\ln z| \le \frac{|\ln(1-\alpha)|}{\alpha}|1-z| \ . \tag{A.1}$$

For $|\ln z| \le \delta$ we have

$$\frac{1-e^{-\delta}}{\delta}|\ln z| \le |1-z| \le \frac{e^{\delta}-1}{\delta}|\ln z|. \tag{A.2}$$

Lemma A.2. *Let $L, H \in \mathcal{L}_+(\mathcal{X})$ such that*

$$\min\left(d_X(L^{-1}, H), d_X(L, H^{-1})\right) \le \delta \ .$$

Then

$$\rho(I - LH) \leq \frac{e^\delta - 1}{\delta} \min\left(d_X(L^{-1}, H), d_X(L, H^{-1})\right) \ . \tag{A.3}$$

Proof. If $\lambda \in \sigma(I - LH)$ then there exists a unit vector $u \in X^{\mathbb{C}}$ such that $(I - LH)u = \lambda u$, therefore

$$(1 - \lambda)u = LHu \ . \tag{A.4}$$

After left-multiplying with L^{-1} and taking the inner product with u we obtain

$$(1 - \lambda)\left\langle L^{-1}u, u\right\rangle = \langle Hu, u\rangle \ , \quad \text{therefore} \quad \lambda = 1 - \frac{\langle Hu, u\rangle}{\langle L^{-1}u, u\rangle} \ .$$

If we substitute $v = H^{-1}u$ in (A.4) and take the inner product with v we have

$$(1 - \lambda)H^{-1}v = Lv \ , \quad \text{therefore } \lambda = 1 - \frac{\langle Lv, v\rangle}{\langle H^{-1}v, v\rangle} \ .$$

Hence, if $d_X(L^{-1}, H) \leq \delta$, then

$$\begin{aligned}\rho(I - LH) &\leq \sup\{|1 - z| \ : \ z = \langle Hu, u\rangle \,/\, \left\langle L^{-1}u, u\right\rangle \text{ for some } u \in X^{\mathbb{C}} \setminus \{0\}\} \\ &\overset{(A.2)}{\leq} \frac{e^\delta - 1}{\delta} d_X(L^{-1}, H) \ .\end{aligned}$$

Instead, if $d_X(L, H^{-1}) \leq \delta$, then

$$\begin{aligned}\rho(I - LH) &\leq \sup\{|1 - z| \ : \ z = \langle Lu, u\rangle \,/\, \left\langle H^{-1}u, u\right\rangle \text{ for some } u \in X^{\mathbb{C}} \setminus \{0\}\} \\ &\overset{(A.2)}{\leq} \frac{e^\delta - 1}{\delta} d_X(L, H^{-1}) \ .\end{aligned}$$

which proves (A.3). ∎

REFERENCES

[1] Alfio Borzi and Volker Schulz, *Multigrid methods for PDE optimization*, SIAM Rev., 51 (2009), pp. 361–395.

[2] Dietrich Braess, *Finite elements*, Cambridge University Press, Cambridge, 1997. Theory, fast solvers, and applications in solid mechanics, Translated from the 1992 German original by Larry L. Schumaker.

[3] Philippe G. Ciarlet, *The finite element method for elliptic problems*, vol. 40 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002. Reprint of the 1978 original [North-Holland, Amsterdam; MR0520174 (58 #25001)].

[4] Juan Carlos de los Reyes, Christian Meyer, and Boris Vexler, *Finite element error analysis for state-constrained optimal control of the Stokes equations*, Control Cybernet., 37 (2008), pp. 251–284.

[5] Andrei Drăgănescu and Todd F. Dupont, *Optimal order multilevel preconditioners for regularized ill-posed problems*, Math. Comp., 77 (2008), pp. 2001–2038.

[6] Andrei Drăgănescu and Cosmin Petra, *Multigrid preconditioning of linear systems for interior point methods applied to a class of box-constrained optimal control problems*, SIAM J. Numer. Anal., 50 (2012), pp. 328–353.

[7] Howard C. Elman, David J. Silvester, and Andrew J. Wathen, *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*, Numerical Mathematics and Scientific Computation, Oxford University Press, New York, 2005.

[8] Vivette Girault and Pierre-Arnaud Raviart, *Finite element methods for Navier-Stokes equations*, vol. 5 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1986. Theory and algorithms.

[9] R. B. Kellogg and J. E. Osborn, *A regularity result for the Stokes problem in a convex polygon*, J. Functional Analysis, 21 (1976), pp. 397–431.

[10] Serge Nicaise and Dieter Sirch, *Optimal control of the Stokes equations: conforming and non-conforming finite element methods under reduced regularity*, Comput. Optim. Appl., 49 (2011), pp. 567–600.

[11] Tyrone Rees and Andrew Wathen, *Preconditioning iterative methods for the optimal control of the Stokes equations*, SIAM J. Sci. Comput., 33 (2011), pp. 2903–2926.

[12] Arnd Rösch and Boris Vexler, *Optimal control of the Stokes equations: a priori error analysis for finite element discretization with postprocessing*, SIAM J. Numer. Anal., 44 (2006), pp. 1903–1920 (electronic).

[13] Rolf Stenberg, *Error analysis of some finite element methods for the Stokes problem*, Math. Comp., 54 (1990), pp. 495–508.